

University of Groningen

Levensecht puzzelen

Jansen, Ritsert C.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2003

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Jansen, R. C. (2003). *Levensecht puzzelen*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Levensecht puzzelen

Rede

*Uitgesproken bij de aanvaarding van het ambt van hoogleraar in de
bioinformatica aan de Rijksuniversiteit Groningen*

op

2 december 2003

door

Dr Ritsert C. Jansen



onder embargo tot dinsdag 2 december 16:15

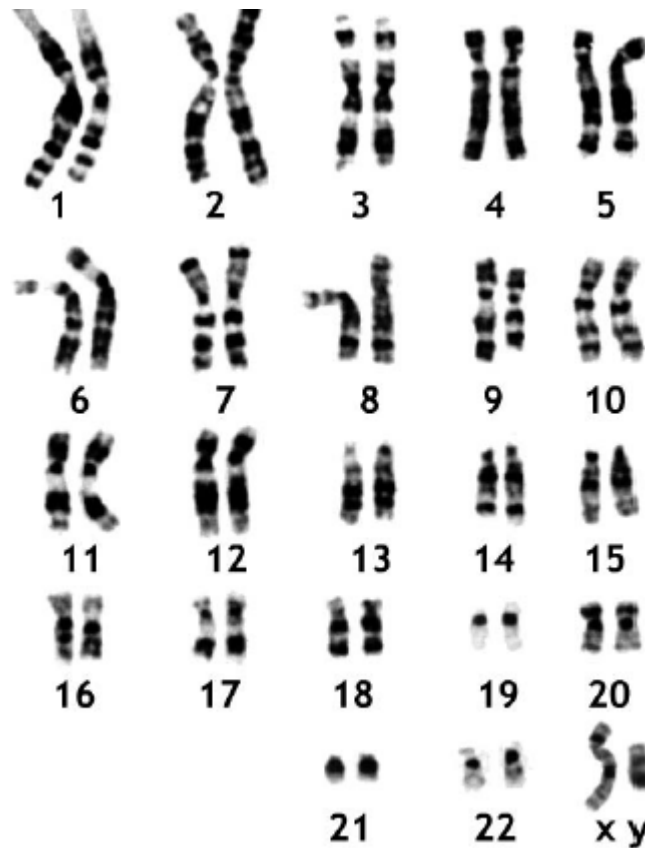
Oratie

Mijnheer de Rector Magnificus,

Dames en Heren,

Puzzelt u wel eens? Waar kijkt u dan als eerste naar als u een puzzel gaat maken? Vast naar hoeveel stukjes die puzzel heeft. Bij 100 stukjes weet u *da's een makkie*, bij 500 krabt u zich al even over de bol, bij 1000 sputtert u tegen *dit gaat mij de hele kerstvakantie kosten*, en bij 10.000 besluit u *geef mijn portie maar een fikkie daar heb ik geen tijd voor*. Wel, het is juist bij dit soort aantallen dat levensecht puzzelen begint. De levenswetenschapper maakt hier – hoe kan het ook anders met zoveel puzzelstukjes – haar levenswerk van. Vandaag zal ik het hebben over de puzzel die de levenswetenschappers willen oplossen en hoe zij daarbij te werk gaan. In het bijzonder gaat mijn aandacht vandaag uit naar de bioinformaticus, die een essentiële rol schijnt te vervullen bij het levensecht puzzelen. Ik, als statistisch geschoold bioinformaticus, ga u in het komende half uur dan ook laten proeven wat deze rol nu inhoudt of in gaat inhouden in Groningen. Wellicht dat enkelen onder u dan de smaak te pakken krijgen. Dat zou mooi zijn, er is meer dan genoeg te puzzelen over.

Levensecht puzzelen. Waarom deze titel? Volgens de “Dikke van Dale” betekent “levensecht” *de werkelijkheid van het leven goed weergevend*. Dit is inderdaad wat we met het eindresultaat van ons gepuzzel beogen te bereiken. Als alle stukjes van de levenspuzzel netjes op hun juiste plaats liggen, geeft de plaat die we dan kunnen aanschouwen de werkelijkheid van het leven goed weer. Bij het woord “puzzelen” staat in de Dikke van Dale de omschrijving *piekeren en peinzen om tot een oplossing te komen*. De woorden piekeren en peinzen zijn mijn inziens bij uitstek van toepassing op de bioinformaticus, die – al piekerend en peinzend – met de door anderen verzamelde informatie puzzelt om het leven zo goed mogelijk te vangen in figuren, schema's en dynamische modellen, om vervolgens de collega's te vertellen hoe het zit, hoe het zou kunnen zitten, of welke puzzelstukjes nog ontbreken. Levensecht puzzelen, dat is dus wat de bioinformaticus zo de hele dag doet – puzzelen als roeping, als hobby, als beroep, want je wordt er nog voor betaald ook! Hoe ver zijn we dan al gekomen met ons gepuzzel en wat staat ons nog te doen? Heel wat, laat daar geen misverstanden over bestaan.

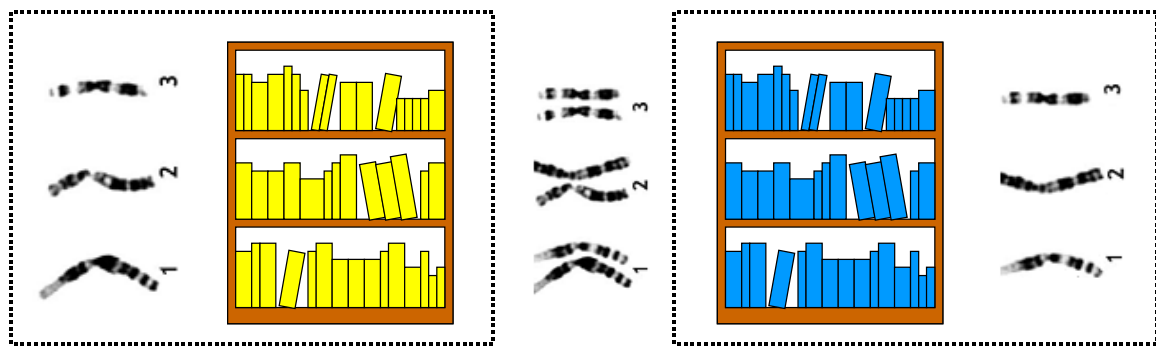


Figuur 1. De 23 paren chromosomen van de mens.

De mens als puzzel

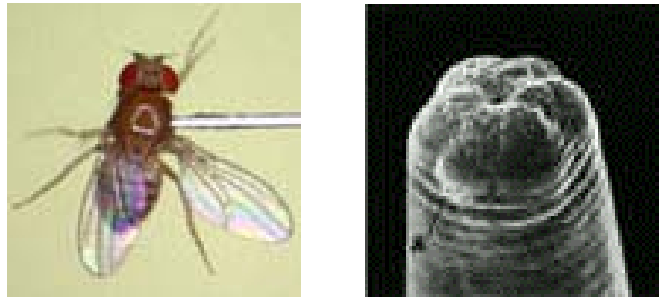
Realiseert u het zich al? U bent een puzzel! Hoe bent u van een bevruchte eicel gegroeid tot wie u nu bent? De blauwdruk hiervoor is vastgelegd in uw erfelijk materiaal, genaamd DNA. Dat DNA is opgeslagen in de kern van uw cellen en is georganiseerd en verdeeld over 23 paren lange draden, de chromosomen. We noemen dit geheel ons *genoom*. U kunt uw DNA voor het gemak beschouwen als een document dat uw bouwplan beschrijft, en ook hoe het hele bouwproces gestuurd moet worden. U bent gewend aan een taal met 26 letters, A tot en met Z. Het DNA, de taal van het leven, is echter geschreven in slechts vier letters, namelijk A, C, T en G. Het hele document van uw bouwplan, een keten van As, Cs, Ts en Gs, is omvangrijk, bij elkaar ongeveer 3 miljard letters. Dat staat gelijk aan de tekst van zo'n 75.000 inaugurele redes. Of, als u bedenkt dat een dicht bedrukt boek zo'n 3000 letters per bladzijde heeft, dan staat het ook gelijk aan 10.000 boeken van 100 bladzijden elk, dat wordt een behoorlijke boekenkast. Opmerkelijk genoeg, elke cel in uw lichaam bevat in de kern zijn eigen boekenkast met een kopie van uw volledige bouwplan. Sterker

nog, het bevat twee bouwplannen, namelijk de versie die u heeft geërfd van uw vader en ook de versie die u van uw moeder heeft ontvangen. Overigens, de tekst in de boeken van uw moeders en vaders kant is vrijwel gelijk. Per bladzijde zijn er 1 à 2 letters verschillend, over alle 10.000 boeken te samen gaat het om zo'n 1.500.000 verschillen.



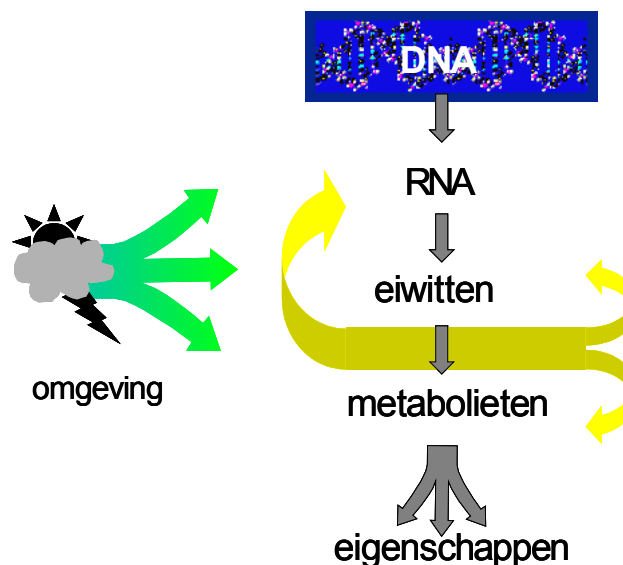
Figuur 2. Elke cel bevat “twee boekenkasten” met het erfelijk materiaal.

Levensrecht puzzelen komt neer op het ontcijferen van het bouwplan van het leven, de taal van ons DNA. We maken vorderingen, we zijn op weg de karakteristieken en betekenis van een groot aantal ‘woorden’ in het DNA te begrijpen. Maar de grammatica, de samenhang tussen de woorden, ontgaat ons nog grotendeels. In het DNA onderscheiden we coderende en niet-coderende stukken en beide soorten DNA bevatten elementen die van belang zijn voor het bouwproces. De grootste aandacht gaat uit naar het coderende DNA, want daar liggen de zogenaamde genen, die coderen voor eiwitten, de bouwstenen en gereedschappen van het leven. Laten we ons dus eerst eens richten op die genen. Wel, hoeveel genen heeft de mens, dus hoeveel puzzelstukjes zijn er? Met de huidige kennis is de schatting dat er ruwweg 30.000 genen zijn. Een flink formaat puzzel niet waar? Hoe staat het met andere organismen? Als vooronderstelling hanteren we natuurlijk dat de mens veel hoger ontwikkeld is dan andere organismen, dus zeker veel meer genen moet hebben dan, pak hem beet, de fruitvlieg *Drosophila melanogaster* of de worm *C. elegans*. Valt dat even tegen, de fruitvlieg heeft 14.000 genen, de worm heeft maar liefst 20.000 genen. Zijn wij nu maar zo'n 1.5 maal complexer zijn dan de worm?



Figuur 3. De fruitvlieg en worm.

Maar laten we ons niet blind staren op louter het *aantal* genen. Laten we nu kijken wat die genen doen. Genen coderen voor eiwitten, de bouwstenen van de cel en de gereedschappen om de cel te laten werken. Genen worden eerst ‘afgelezen’ en het aldaar gecodeerde woord wordt gekopieerd en de kopie, die we RNA noemen, wordt de cel ingestuurd. In de cel worden de RNA moleculen weer vertaald naar de eiwitten. Elk gen codeert voor een of meer specifieke eiwitten. Een gen kan meer of minder aan staan, en dus meer of minder RNA en eiwit produceren, afhankelijk van de omstandigheden. Deze eiwitten spelen weer een belangrijke rol in de stofwisseling, en regelen de aanmaak en hoeveelheid van allerlei voor de levensprocessen nuttige tussen- en eindproducten – metabolieten genaamd. Overigens, als we de buzzwords genomics, transcriptomics, proteomics en metabolomics in de mond nemen, dan doelen we gewoon op grootschalig DNA, RNA, eiwitten en metabolieten meten.



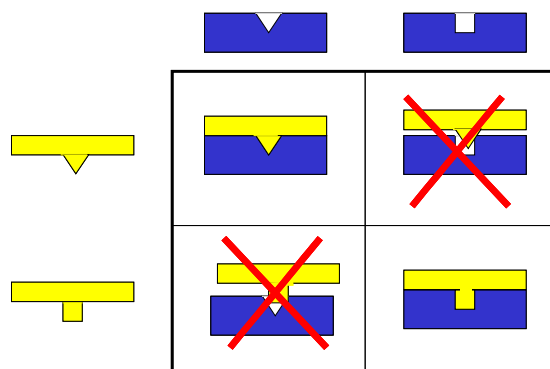
Figuur 4. Het centrale dogma in de biologie.



Figuur 5. Een demonstratie van paarsgewijze interacties.

De 30.000 genen werken met elkaar samen doordat hun gen-producten elkaar in de cel gaan ontmoeten en aan elkaar of aan het DNA passen. In een gewone puzzel past elk puzzelstukje op een beperkt aantal andere puzzelstukjes: vorm en beeld moeten overeenkomen. Daarentegen in de levenspuzzel weten we niet van te voren met hoeveel andere puzzelstukjes een gegeven stukje zal combineren of – beter gezegd – interacteren. In potentie kan je één interactie hebben tussen twee genen, drie paarsgewijze interacties en een drievoudige interactie tussen drie genen, en ga zo maar door. Een goed voorbeeld van paarsgewijze interacties ontstaat als mensen elkaar de hand geven – om u dit te demonstreren heb ik even de hulp in geroepen van drie in Groningen gerenommeerde acteurs: op de foto ziet u hoe de drie heren de drie paarsgewijze interacties netjes demonstreren - ik dank de heren hartelijk voor hun bijdrage aan mijn oratie. Al deze mogelijke interacties vergroten de complexiteit van de puzzel natuurlijk enorm. Maar er is nog een andere dimensie aan het begrip interactie, en die wil ik u ook uitleggen aan de hand van het voorbeeld van handen schudden. Als we elkaar de hand schudden, dan gaat dat functioneel prima zolang we maar dezelfde hand gebruiken. Steekt de één echter zijn of haar rechterhand uit, en de ander de linkerhand, dan blijkt het proces van handen schudden ineens veel lastiger. Afijn, ook een gen kan in twee of meer varianten of toestanden voorkomen, net als links- en rechtshandigheid, ten gevolge van veranderingen in de DNA codering of afschrijving. De twee of meer varianten van een gen coderen voor twee of meer varianten van de bouwstenen en gereedschappen, en die verschillende varianten kunnen dus bij wijze van spreken ook weer rechts- en linkshandig zijn – in de figuur

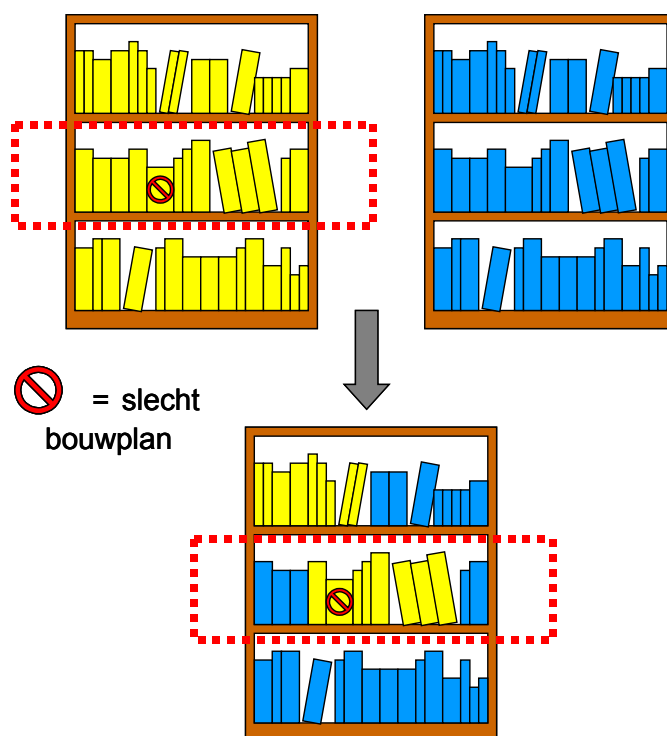
ziet u een voorbeeld. Een ligand en een receptor komen elk in twee varianten voor, waarbij de ligand en receptor slechts in twee van de vier combinaties goed op elkaar passen. Hierin schuilt een orde van complexiteit die we nog niet hebben beschouwd. Stel dat elk gen twee varianten of toestanden kent, dan zijn er bij de mens in totaal $2^{30.000}$ verschillende combinaties mogelijk – geen wonder dat elk mens anders is. Het equivalent bij de worm met zijn 20.000 genen is dan slechts $2^{20.000}$. Zo gezien is de mens een factor $2^{30.000}$ gedeeld door $2^{20.000}$ complexer dan de worm, een slordige 1 met 3000 nullen er achter. Kortom, de menselijke puzzel is mogelijk heel wat complexer dan die van de worm. Zo gezien, gaat de parallel 30.000 genen = 30.000 puzzelstukjes ook niet echt op, de mensenpuzzel is veel complexer omdat ieder stukje op veel andere stukjes past en vooral ook omdat verschillende mensen verschillende puzzelstukjes kunnen hebben die per mens weer anders kunnen combineren. De puzzel is dus moeilijker op te lossen ten gevolge van de vele complexe interactiepatronen. In werkelijkheid is het aantal interacties wel minder dan hier aangegeven, maar zelfs als we er een factor miljoen naast zitten is de complexiteit van de levenspuzzel nog schrikbarend hoog.



Figuur 6. Een voorbeeld van interacties tussen gen-varianten.

Toch was dit nog niet alles. De levenspuzzel is niet simpelweg tweedimensionaal (zeg maar plat op tafel) maar driedimensionaal (zulke puzzels waren ook een tijdje populair in de speelgoedwinkels), de puzzel heeft deeloplossingen per orgaan of weefseltype (het bouwplan bevindt zich in elke cel, maar gelukkig gaan alleen de juiste genen op de juiste plaats aan), de puzzelstukjes kunnen veranderen in de tijd (u zult niet verbaasd zijn dat elke levensfase zijn eigen patroon van genactiviteit kent), en tenslotte, de puzzel wordt beïnvloed door externe factoren (of u gezond of ongezond leeft, bijvoorbeeld). Zo gezien is het wonder, nee, bent u een wonder, dat al

die processen over het algemeen toch zo goed en haast als van zelf functioneren. Maar het gaat toch ook vaak mis. Klassieke voorbeelden zijn kystische fibrose (taaislijmziekte) en de ziekte van Huntington, een ziekte die bepaalde delen van de hersenen aantast. Deze ziekten worden elk grotendeels bepaald door een probleem in één gen, zo heeft onderzoek uitgewezen. In hoekjes van de levenspuzzel boeken we dus deelsuccessen. De wetenschappelijke uitdaging ligt er natuurlijk in om de levenspuzzel verder af te maken ook in die delen van de puzzel die gaan over de complexere ziekten zoals kanker of diabetes, waar aanleg en omgeving van belang zijn. Om bij te dragen aan de kwaliteit van het leven, om gezonder oud te worden.

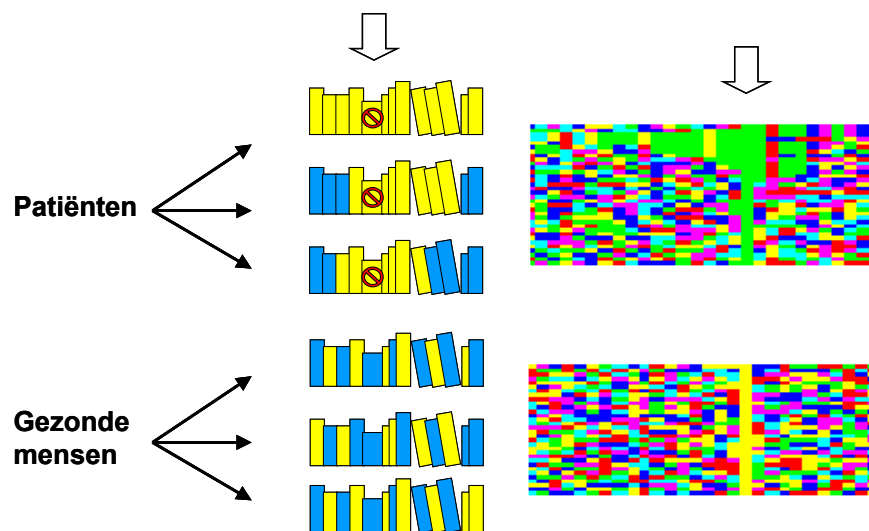


Figuur 7. Overerving van erfelijk materiaal van ouder op kind.

In 2001 is de eerste versie van het hele menselijke genoom gepubliceerd in Science en Nature, dat wil zeggen, voor het eerst is van één mens de hele DNA volgorde afgelezen en in de computer opgeslagen. Let wel het aflezen van de hele DNA volgorde van meerdere mensen is nog een flink aantal stationnetjes verder. Op dit moment is het dus nog niet mogelijk om *uw* boekenkast te vullen met *uw* volledige DNA code. Wel is het mogelijk om uw DNA te karakteriseren op een groot aantal plaatsen voor de aldaar aanwezige DNA varianten, met behulp van zogenaamde moleculaire merkers. We spreken ook wel over een *genetische vingerafdruk*. In de terminologie van de boekenkast, bestaat uw genoom dus uit zo'n 10.000 boeken die u

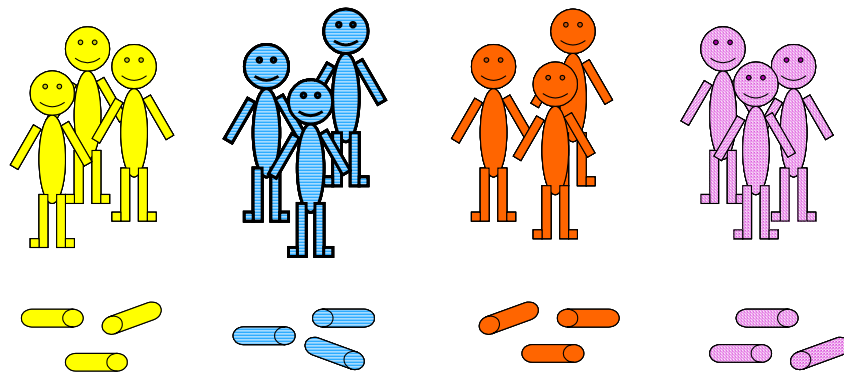
van uw vader hebt geërfd, en 10.000 boeken afkomstig van uw moeder. De genetische vingerafdruk vertelt ons nu louter welke kleuren de ruggen van uw boeken hebben – de rug van een boek zien we zonder dat we het boek verder openslaan om het te lezen. Een genetische vingerafdruk kan een beperkt beeld geven, zeg de kleur van de ruggen van zo'n 100 verschillende boeken uit uw boekenkast, of - als u meer geld heeft - een meer compleet beeld, dus voor alle 10.000 boeken in uw boekenkast. Met zo'n vingerafdruk kunnen we vervolgens de overerving van ons erfelijk materiaal van ouder op kind volgen. Hoe gaat dat globaal in zijn werk? De boeken die u van uw vader heeft geërfd, hebben bijvoorbeeld een gele rug, terwijl de boeken afkomstig van uw moeder een blauwe rug hebben. Aan uw kinderen geef u een complete set van boeken door, waarbij u soms een hele plank met gele boeken doorgeeft, dan weer een plank met blauwe boeken – dit heet segregatie van chromosomen in genetische termen, een plank staat voor een heel chromosoom. U kunt ook een *deel* van een plank met gele boeken doorgeven, en dit dan aanvullen met blauwe boeken, of vice versa – zodat uw kind toch weer een compleet bouwplan heeft – dit heet overkruising of recombinatie tussen chromosomen in genetische termen. Uw kind heeft dus een mengelmoes van uw vaders en moeders boekenkast via u ontvangen – een mengelmoes van uw vaderlijke en moederlijke genomen. Elk kind van u krijgt een andere mengelmoes van de gele en blauwe boeken, want de processen van segregatie en recombinatie worden geregeerd door het toeval. Als nu in uw vaders bouwplan ergens in een bepaald boek een slecht ontwerp voor een bepaald orgaan of proces staat, dan zal uw kind dat erven als u hem of haar dat betreffende boek per ongeluk meegeeft. In een stamboom met meerdere individuen kunt u dan zien dat een ziektebeeld – een slecht ontwerp - steeds samenvalt met het hebben van dat bepaalde boek met die gele rug en doorgaans met nog enkele andere boeken met geel kft die naast dat gele boek staan, want er worden altijd rijtjes boeken doorgegeven van ouder op kind – zo'n rijtje boeken noemen we een *haplotype*. Het duurt een behoorlijk aantal generaties voordat dat rijtje boeken kleiner wordt ten gevolge van recombinaties. Dit gegeven kunnen we benutten als we in een populatie, zeg die van Noord-Nederland, de haplotypes van patiënten en gezonde mensen vergelijken. Patiënten hebben het ziektegen van een gemeenschappelijke voorouder gekregen, en zij hebben doorgaans in de buurt van het boek met het ziektegen nog een rijtje met gelijk gekleurde boeken van die gemeenschappelijke voorouder staan, als er tenminste niet te veel generaties overheen zijn gegaan. Dat rijtje met gelijk gekleurde boeken is

dan langer bij patiënten dan bij willekeurige gezonde personen - een gegeven dat door Gerard te Meerman en collega's in Groningen is benut om een methode voor de analyse van humane populaties te ontwikkelen – op basis van *haplotype sharing*. Op deze wijze kunnen we één gen tegelijk ontdekken, en er zijn aanvullende methoden ontwikkeld om te testen of er in de nabijheid nog andere ziektegenen liggen. Voor complexere eigenschappen en processen, waarbij meerdere genen betrokken zijn, vergt deze onderzoeksroute echter enorm grote populaties en de beschikbaarheid van een voldoende gedetailleerde genetische vingerafdruk – wij werken momenteel samen met Galileo Genomics, een hightech bedrijf in Canada, waarmee wij deze grootschalige analyses verwachten te realiseren, om aldus een beter beeld te krijgen van complexe ziekten als astma en kanker.



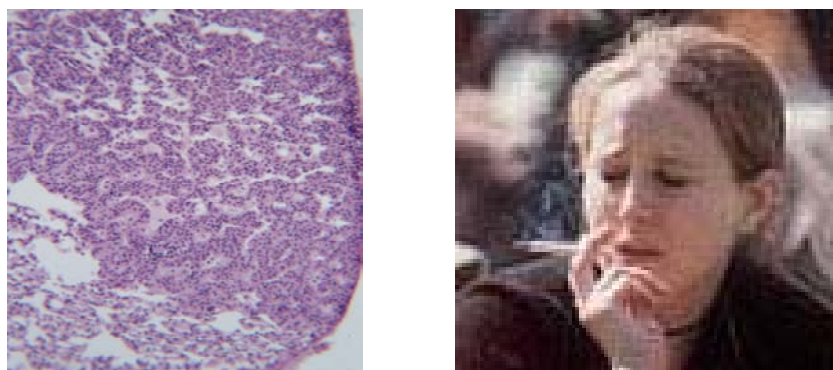
Figuur 8. Patiënten hebben vaker dezelfde boeken in de buurt van het boek met het ziektegen staan (rechts voor een groter aantal individuen).

Een belangrijke vraag is of en hoe we zinvolle uitspraken over preventie en behandeling kunnen doen op basis van dit soort grote populatiestudies. Leest u wel eens de bijsluiters van medicijnen, dan weet u dat ze vol waarschuwingen staan over mogelijk bijwerkingen, waar de één wel en de ander geen last van krijgt. Hadden we dan geen beter medicijn kunnen ontwikkelen? De moeilijkheid is dat ieder mens feitelijk een unieke meervoudige combinatie van gen-varianten is. Wat we leren over wat goed is voor mij, hoeft niet voor u te gelden, en vice versa. Wij zijn dus niet klaar met levensecht puzzelen voordat we maatwerk voor individuen of specifieke groepen kunnen leveren. Dit vereist kennis over de interacties tussen genen, en over de interacties tussen genen en medicijnen. Leveren van maatwerk in preventie en



Figuur 9. Medicijn op maat.

interventie, dat is het werkteerrein van de *pharmacogenetics* en *pharmacogenomics*. Een mooie combinatie van genetisch en genomisch onderzoek, van grote economische betekenis voor de farmaceutische industrie. Maar dit is niet alles. In het geval van bijvoorbeeld longkanker kunnen we wel veel investeren in genetisch en genomisch onderzoek, en dat kan tot relevante kennis leiden, maar een algemeen rookverbod zal directer tot een afname van het aantal gevallen van longkanker leiden en zal dus een veel grotere bijdrage aan de volksgezondheid leveren. Evenzo spelen slechte voedingsgewoonten en te weinig beweging ons parten. Ouderdomsdiabetes is hiervan een goed voorbeeld. We zijn dus niet klaar met levensecht puzzelen voordat we de invloed van alle betrokken genetische en niet-genetische factoren tegen elkaar hebben afgewogen – op basis van goed gefundeerd epidemiologisch, genetisch en bioinformatica-onderzoek.



Figuur 10. Roken leidt tot longkanker.

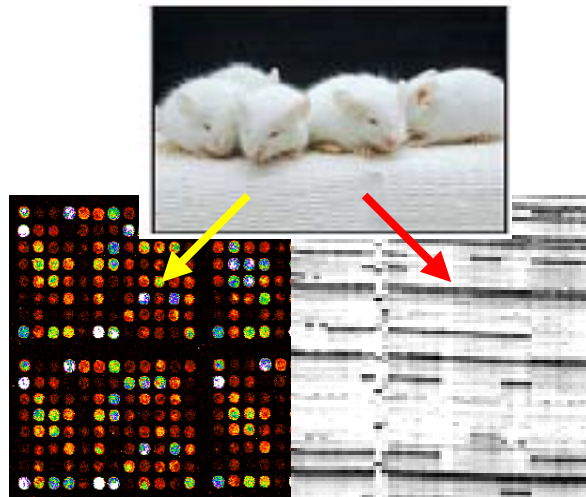
Bij de mens is het mogelijk om biomoleculaire data te verzamelen, zij het toch op relatief beperkte schaal. Uw DNA kunnen we extraheren uit bijvoorbeeld een haar of

bloed. Meten van RNA, eiwit en metabolieten in lichaamsvloeistoffen als speeksel en bloed gaat ook prima, u zult het echter niet waarderen als uw dokter u een stukje hersen- of leverweefsel wil afnemen omdat hij uw RNA, eiwit of metabolietenprofiel ter plaatse wil vaststellen. Dit soort weefsel komt wellicht beschikbaar in geval van een noodzakelijk geachte operatie van een patiënt, maar nooit van gezonde personen. De levenswetenschapper kan dus niet zien welke genen zich anders gedragen bij patiënten dan bij gezonde mensen. De ethiek stelt hier terecht grenzen aan wat mogelijk is bij de bestudering van de mens. Willen wij dan ook levensecht puzzelen in al zijn dimensies, dan ontkomen we er niet aan om met diermodellen te werken – waarbij de ethische grenzen ook weer scherp bewaakt dienen te worden.

De muis als puzzel

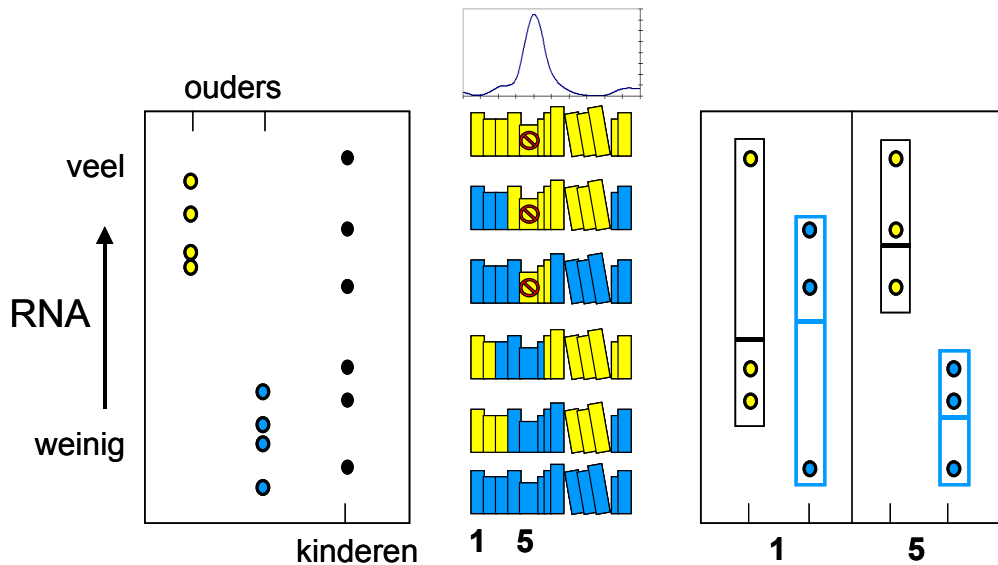
Levensecht puzzelen aan de mens kent dus zijn beperkingen. Om sneller vooruitgang te boeken met het oplossen van alle dimensies van de levenspuzzel, dienen we naar mijn mening het onderzoek aan de mens te complementeren met gericht onderzoek aan proefdieren. Laten we nu eens kijken wat proefdieren kunnen bijdragen, neem de huismuis. Het bouwplan van de huismuis lijkt sterk op dat van de mens. Het DNA van de muis is beschreven in een taal die we als een dialect van onze DNA taal kunnen beschouwen, en met levensecht puzzelen aan de muis leren we een hoop over de mens. Helaas voor de huismuis, worden bij dit dier ongevraagd wel RNA-, eiwit- en metabolietprofielen aan allerlei lichaamsweefsels vastgesteld, naast allerlei andere klinisch interessante parameters, bij zowel gezonde als ziekgemaakte individuen. Om met een stuk uit de New York Times van december 2002 te spreken *de huismuis leeft onuitgenodigd in onze huizen en steelt ons voedsel, echter de huismuis is nu haar schuld dik aan het terugbetalen*. Gepaste eerbied voor de huismuis is dus op haar plaats. Evenals een kritische houding ten aanzien van het gebruik van proefdieren. Naast de muis dient een heel scala aan andere organismen als model, variërend van relatief eenvoudige organismen waar de DNA-boekenkast relatief klein is, zoals de eerder genoemde worm *C. elegans* of het bakkersgist *S. cerevisiae*, tot meer complexe organismen als de fruitvlieg *Drosophila melanogaster* en de rat. Onderzoek aan modelsystemen heeft een hoge vlucht genomen aan vooraanstaande universiteiten in binnen- en buitenland. Ik kan u melden dat wij in Groningen in de startblokken staan

om in de muizenwereld een duidelijke positie in te gaan nemen via een internationaal consortium – het Complex Trait Consortium. In 2004 vindt de jaarlijkse conferentie van dit consortium plaats op het fameuze Jackson Laboratory in de Verenigde Staten, in 2005 zal die bijeenkomst door ons in Groningen georganiseerd worden – voorwaar iets om trots op te zijn.



Figuur 11. RNA expressie (links) en genetische vingerafdruk (rechts).

Wat kan er bij de muis? Net als bij de mens, kunnen we ook bij de muis naar de ouder – kind relaties kijken, en de overerving van de boeken in de boekenkast volgen via de ‘gele’ en ‘blauwe’ kaften. De effecten van de genen kun je het beste bestuderen als de ouders flink van elkaar verschillen en als ze samen veel kinderen krijgen. We kiezen nu twee muizenstammen, zeg één stam die extreem gevoelig is voor kanker met gele boeken in de boekenkast, en een andere stam die relatief ongevoelig is met blauwe boeken in de kast. En net als bij de mens, gaan we de wisselende samenstelling van gele en blauwe boeken bij de nakomelingen bestuderen. Liefst bij veel nakomelingen, en dat is bij de muis goed te realiseren. De vraag is dan hoeveel genen er betrokken zijn bij de eigenschap kanker, waar die genen op het genoom liggen, wat hun effecten zijn, en hoe ze “met elkaar praten”. Bij de muis kunnen we, zoals gezegd, de levenspuzzel grootschalig bestuderen op het niveau van RNA, eiwitten en metabolieten. De volgende stap in de bioinformatica aanpak ligt nu erg voor de hand: voor elk RNA, elk eiwit of elke metaboliet gaan we op zoek naar die boeken in de boekenkast, waarbij de nakomelingen met het gele kaft altijd meer RNA, eiwit of metaboliet hebben dan de nakomelingen met het blauwe kaft, of vice versa. Vinden

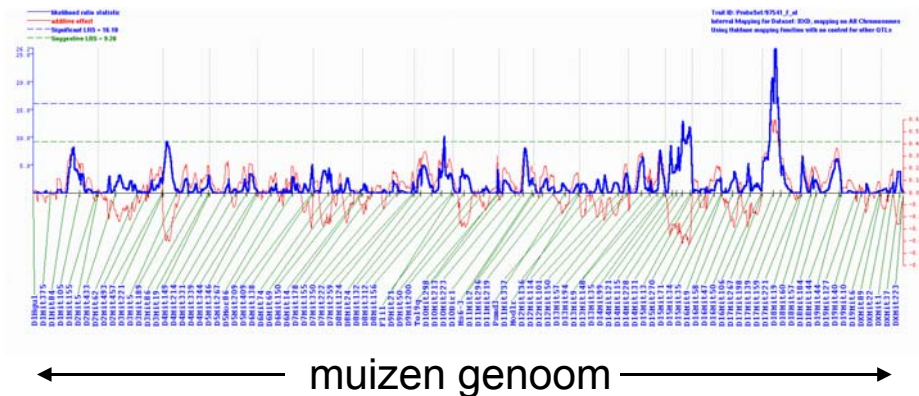


Figuur 12. Een spoedcursus 'genetical genomics'.

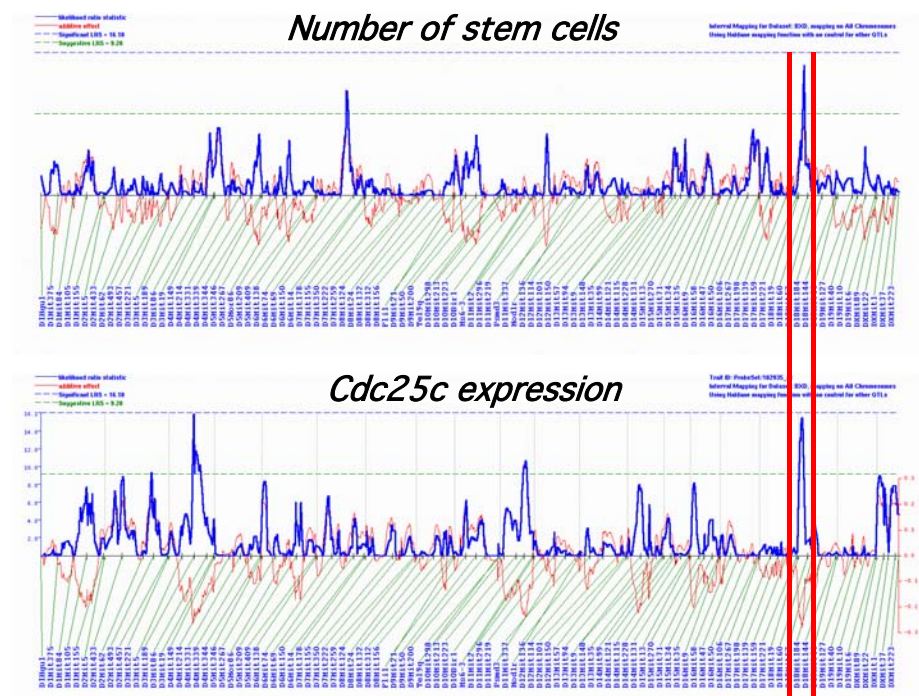
we één of meer van zulke boeken, dan weten dat de hoeveelheid RNA, eiwit of metaboliet gereguleerd wordt door een ander gen, en we weten dan ook bij benadering waar dat andere gen ligt. Laat ik u nu uitleggen hoe dat in zijn werk gaat – u krijgt nu een spoedcursus in de detectie en kartering van genen. Ik laat u nu een illustratie zien die gebaseerd is op een experiment met *bakkersgist*, gepubliceerd in Science in 2002 door een Amerikaanse groep, om u het principe uit te leggen. Voor een bepaald gen is de RNA hoeveelheid een aantal keer gemeten voor elk van beide ouders, en u kunt zien dat de ene ouder (aangegeven met geel) duidelijk hogere waarden heeft, en de andere ouder (aangegeven met blauw) lagere waarden – hoeveel genen zouden hieronder liggen en waar zouden die genen op de chromosomen liggen? De zes nakomelingen in dit figuur geven RNA waarden van laag tot hoog te zien. Hoe kunnen we deze informatie nu gebruiken om het regulerende gen of de regulerende genen te vinden? Op de schematische figuur in het midden ziet u, in boekenkasttaal, de boekenplanken van de zes nakomelingen. U ziet dat elke nakomeling een andere combinatie van gele en blauwe boeken heeft op deze boekenplank. In het rechter figuur ziet u op de positie aangegeven met nummer 1 dat de individuen met gele en blauwe kaften zowel lage als hoge RNA waarden. Gaan we dan naar positie 5, dan zien we dat de nakomelingen met een geel kaff een hoge RNA waarde hebben, en met een blauw kaff een duidelijk veel lagere waarde. Dit is een sterke indicatie dat de RNA hoeveelheid bepaald wordt door een gen dat beschreven staat in boek 5 op deze plank en dat twee varianten heeft, een gele en een blauwe variant. In de figuur

midden-boven geeft de piek de meest waarschijnlijke positie aan. Dit is een illustratie van een basale analyse aan één gen. Onze levenspuzzel heeft echter vele stukjes, de 30.000 genen zijn goed voor tenminste 30.000 verschillende RNAs, en dus voeren we zo'n 30.000 van dit soort analyses alleen al op RNA niveau uit (om nog maar te zwijgen over de eiwitten en metabolieten). Het interpreteren van alle resultaten van al die analyses, dat wordt dus een hele klus en het voert te ver om dat hier te gaan uitleggen. Deze strategie hebben dr Jan-Peter Nap en ik in 2001 in het tijdschrift Trends in Genetics gepubliceerd onder de naam 'genetical genomics'. Als reclame voor de methode voerden we aan dat het een helemaal nieuwe en krachtige manier zou zijn om gen-netwerken te (re)construeren en kandidaat-genen aan te wijzen. Door gebruik te maken van genetische variatie, geeft genetical genomics inzicht in de causale verbanden in gen-regulatie en geeft concrete getallen voor de grootte van de effecten die de genen op elkaar kunnen hebben. Veronderstelde causale verbanden kunnen vervolgens met gerichte experimenten worden gecontroleerd. RuG-collega dr Gerald de Haan is in 2001 samen met Amerikaanse collega's 'genetical genomics' experimenten gaan opzetten met muizen, waarbij in Groningen de aandacht uitging naar eigenschappen van stamcellen, en in de Verenigde Staten naar hersencellen. Een Amerikaans bedrijf heeft de RNA metingen voor haar rekening genomen. Zulk werk kost een boel inspanning, tijd en geld, maar het afgelopen half jaar is een groot deel van de gegevens binnengekomen. Ik prijs mij dan ook heel gelukkig dat ik bij mijn overgang naar Groningen een gespreid 'genetical genomics' bedje aantrof – uniek in Nederland en daarbuiten. Hoewel er nog veel te doen is over hoe deze gegevens precies geanalyseerd zouden moeten worden, wil ik u toch twee plaatjes van eerste voorlopige analyses niet onthouden. *The proof of the pudding is in the eating* – en dat geldt met name voor de bioinformaticus. Theorie is mooi, maar het moet wel werken. Het eerste voorbeeld betreft het gen met de naam H2-D. We kennen de locatie van dit gen, en weten dat de twee ouderstammen verschillende varianten met verschillende RNA hoeveelheden hebben. Bij analyse zou dit gen dus naar zichzelf moeten wijzen als veroorzaker van de RNA variatie – en dit plaatje laat zien dat dat het geval is: de piek aan de rechter kant geeft de meest waarschijnlijke positie aan op het genoom en hier ligt het H2-D gen inderdaad. Een tweede voorbeeld laat zien hoe je de verschillende puzzelstukjes aan elkaar kunt knopen. Gerald de Haan en zijn collega's hebben de hoeveelheid RNA van het gen met de naam Cdc25c gemeten en aan

H2-D-locus



Figuur 13. Proof of the pudding is in the eating.



Figuur 14. Ligt tussen de rode lijnen een gemeenschappelijke regulator?

dezelfde muizen ook het aantal stamcellen geteld. Het is bekend uit ander onderzoek dat Cdc25c inderdaad betrokken is bij productie van stamcellen. Hoe passen de puzzelstukjes nu op elkaar? Zoals u in de figuur kunt zien, zowel Cdc25c als ook het aantal stamcellen wordt beïnvloed door een gen dat ligt tussen de aangegeven rode lijnen. Weer twee stukjes in de levenspuzzel die netjes op elkaar passen! Prof Rob Williams, een Amerikaanse collega van Gerald de Haan, werkt met dezelfde muizen. Hij verzamelt echter geen gegevens over stamcellen, dan doen we hier in Groningen

al, hij verzamelt gegevens over processen in hersencellen. Het internationale Complex Trait Consortium maakt het zelfs mogelijk dat vele andere onderzoekers met heel diverse interesses aan dezelfde muizenstammen onderzoek kunnen gaan doen, en vanuit Groningen zullen we ook de integratie van al deze informatie ter hand nemen. Het mooie is nu dat we deze heel verschillende informatiebronnen aan elkaar kunnen gaan knopen in de komende tijd – op dezelfde wijze als we net aantal stamcellen gecorreleerd hebben aan de expressie van het gen *Cdc25c*. Combineren we dit met informatie over nog meer weefsels en organen, dan ontstaat hieruit er op termijn een atlas van de variatie en regulatie van genexpressie bij de muis. Zo'n atlas vraagt om goede visualisatie, en deze activiteit zou dan ook een mooie samenwerking met visualisatiecollega prof Jos Roerdink kunnen opleveren. Hoe het ook zij, de komende tijd zullen de bioinformatici Gerald de Haan graag bijstaan met puzzelen aan zijn gegevens.

De huismuis is een modelsysteem, we zijn niet echt in de muis zelf geïnteresseerd, maar wel in de brug die we er mee kunnen slaan naar de mens. Bacteriën, zoals *Lactococcus lactis* en *Bacillus subtilis*, met elk slechts een paar duizenden genen, kunnen ook gezien worden als modelsystemen, en Groningen neemt met haar onderzoek aan bacteriën internationaal een sterke positie in. Met de bacterie kan men geen gecontroleerde kruisingen maken zoals dat bij de muis kan, maar een 'genetical genomics' strategie op basis van natuurlijke genetische diversiteit behoort tot de mogelijkheden; om zo de interacties tussen gen-varianten en tussen gen-varianten en omgevingsfactoren te bestuderen. Aan zo'n project zouden wij graag meewerken. Wat wel is opgestart, en waarmee ik mij heel gelukkig prijs, is een project van en met RuG-collega prof Oscar Kuipers. In dit project wordt met één en dezelfde bacteriestam gewerkt, er zijn dus geen genetisch verschillende individuen in dit project. De mogelijke interacties tussen stukjes in de levenspuzzel worden in dit project bestudeerd door in vitro te testen welke eiwitten wel en niet onderling kunnen interacteren, te kijken of deze eiwitten inderdaad in levende bacteriënlijke tegelijkertijd op dezelfde plaats in de cel voorkomen, en door te kijken of de RNA hoeveelheden van deze eiwitten over verschillende groeistadia een gelijk patroon vertonen. Een uitdagend project! De bacterie is overigens niet louter een mooi modelsysteem. Onderzoek aan de bacterie is hoogst relevant, want bacteriën zijn natuurlijk enerzijds potentiële ziekteverwekkers en anderzijds nuttige spelers in de

voedingsindustrie.

Ook de plantenwereld heeft een model systeem, namelijk *Arabidopsis thaliana*, of wel de zandraket. In samenwerking met Wageningen en Utrecht zullen we binnen een half jaar de beschikking hebben over ‘genetical genomics’ gegevens in de model plant *Arabidopsis thaliana*, of wel de zandraket. Wij kijken in Groningen met spanning uit naar deze gegevens en de biologische implicaties van de resultaten van de bioinformatica analyses! Met *Arabidopsis thaliana* wordt een brug geslagen naar lastiger te onderzoeken maar commercieel belangrijke gewassen.



Figuur 15. Zandraket, bakkersgist, rat en bacillus

Puzzelen met wiskunde, statistiek en informatica

Een reëel gevaar, zeker bij grote “omics” initiatieven, is dat de laboratoriumonderzoekers nauwelijks verder hebben gedacht dan dat de gegevens verzameld moeten worden. De waarde van de verzamelde gegevens blijkt immers achteraf wel – desnoods gaan we blind aan *datamining* doen om het beloofde goud te vinden. Wat en hoe die gegevens dan opgeslagen en geanalyseerd zouden moeten worden, tsja, dat komt later wel, eerst de gegevens maar verzamelen. Overleg met de bioinformaticus vooraf is echter cruciaal, om te voorkomen dat *de patiënt in feite al overleden is voor hij op de operatietafel van de bioinformaticus is beland*. Of kunt u zonder bioinformaticus? U heeft gehoord over de grote boekenkast met de 10.000 boeken van 100 bladzijden elk. In die boeken staat de informatie over de 30.000 genen, zo’n 3 genen per boek, ongeveer een bladzijde per gen. Wat denkt u, gaat u zo’n boek eens lekker in bed liggen lezen? Dat is toch geen bezigheid voor een mens, al die As, Cs, Ts en Gs, wat moet je er mee? En alsof het DNA ons al niet genoeg uitdaging biedt, verzamelen we ons vervolgens ook nog suf aan al het andere wat los en vast zit, van RNA, eiwitten en metabolieten tot klinische, fysiologische of anderszins relevante parameters. Waar en hoe zullen we al die informatie eens

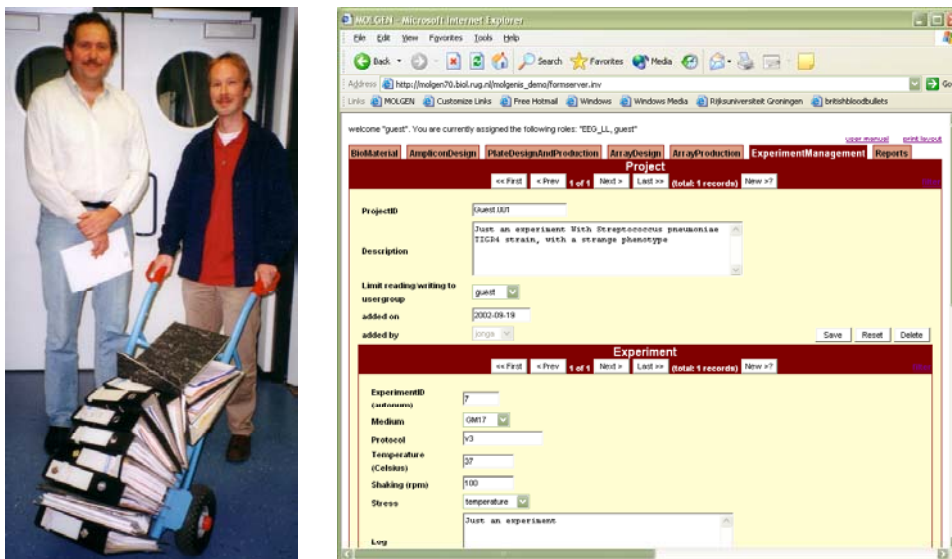
opbergen? En dan, wat moeten we dan met die brei aan gegevens? U begrijpt, de handen van de levenswetenschapper gaan in het haar, zo wordt het nooit wat met levensecht puzzelen, ware het niet dat de bioinformaticus hier als reddende engel te hulp komt schieten. De bioinformaticus, die heeft een koffertje bij zich met daarin diverse cruciale puzzelgereedschappen uit de wiskunde, statistiek en informatica. Informatie management systemen, wiskundige modellen, rekenalgoritmes, parallel rekenende computers, enzovoorts. De bioinformaticus is daardoor in staat tot maatwerk en stelt nieuwe gereedschappen samen al naar gelang de behoefte bij het puzzelen. Voor de duidelijkheid, ook de bioinformaticus kan geen ijzer met handen breken, er zijn dus grenzen aan zijn of haar bijdrage.



Figuur 16. De 1^e dinsdag in december.

Puzzelen met 30.000 stukjes, hoe voorkomt u eigenlijk dat u onderweg stukjes of de kluts kwijt raakt? Bijvoorbeeld, is een onderzoeker vertrokken, dan zijn de gegevens weg of althans niet meer te doorgronden. Of, u wilt de gegevens over hersencellen combineren met de gegevens over stamcellen, maar er ligt een oceaan tussen, want het onderzoek aan stamcellen vindt plaats in Nederland, en dat aan hersencellen in Amerika. Of, u heeft tussenresultaten van een analyse, en leent een steekkarretje om ze naar uw klant te brengen? Het is duidelijk, een high-throughput “omics” laboratorium kan vandaag de dag niet zonder adequaat informatie systeem gebaseerd op moderne Informatie en Communicatie Technologie. Zo’n systeem dient aan allerlei eisen te voldoen, waarbij het kunnen meegroeien met de soms stormachtige technologische ontwikkelingen, niet de minste eis is. Drs Morris Swertz en dr Bert de Brock hebben Groningen Bioinformatica Centrum een niche in de markt geleverd onder de naam MOLGENIS – molecular genetics information system – waarbij we zeer verheugd zijn dat we recent samenwerkingspartners hebben geworven bij laboratoria gericht op de bacterie, plant, dier en mens. MOLGENIS richt zich nu op

de verschillende DNA chip technieken zoals microarrays en Affymetrix chips, en zal ook ons startpunt vormen voor het proteomics lab en mogelijk ook voor het genotyperingslab. We voorzien diverse additionele ontwikkelingen, zo zal MOLGENIS ook een vorm van Statistische Proces en Kwaliteitscontrole bieden. En last-but-not-least, we zullen MOLGENIS als open-source product op de markt brengen – als één van de visitekaartjes van Groningen Bioinformatica Centrum.



Figuur 17. Data management. Fout (links) en goed (rechts).

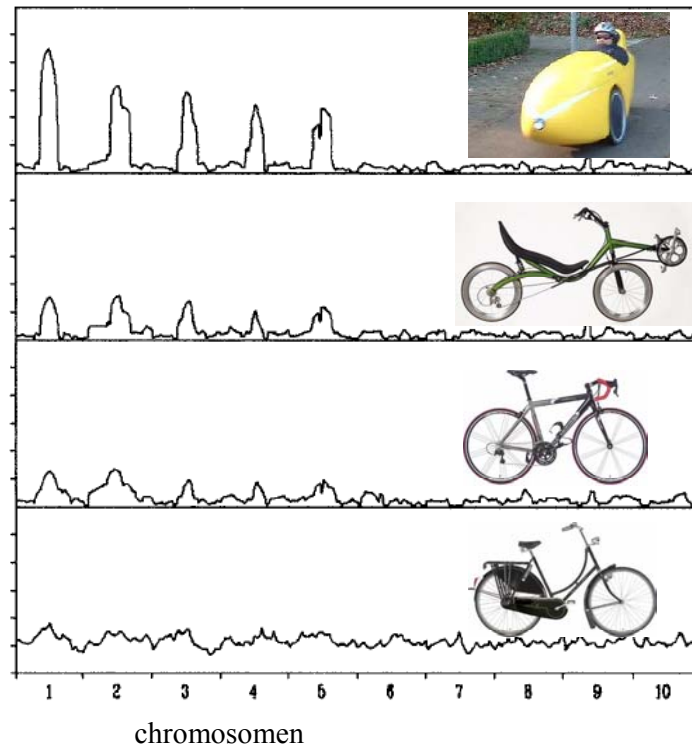
U puzzelt met 30.000 of meer stukjes, hoe pakt u dit zo efficiënt en effectief mogelijk aan? We hebben gezien, het leven is complex en multifactorieel bepaald – genetica, dieet, infecties, milieu, gedrag, sociale structuren, het heeft allemaal invloed. Hoe kunnen we de levenspuzzel te lijf gaan? Zit er nog iets nuttigs in de koffer van de bioinformaticus dat ons hier kan helpen? Ja zeker, heel veel nuttigs zelfs: bijvoorbeeld een ruim arsenaal aan statistische gereedschappen bedoeld voor het ontwerpen van efficiënte en effectieve multifactoriele experimenten, voor het modelleren van het samenspel van al deze factoren op biologisch relevante wijze, en voor het analyseren van de gegevens om na te gaan welke factoren belangrijk zijn en welke niet! De levenswetenschapper heeft de neiging om een complexe eigenschap stapje voor stapje te gaan bestuderen – het is al complex genoeg - dus door bijvoorbeeld steeds maar één gen tegelijk uit te schakelen en dan te kijken wat er verandert. Hoewel het een simpele aanpak is, begint het besef te komen dat het niet de meest effectieve en meest economische strategie is. De al eerder genoemde Rob Williams, directeur van het Complex Trait Consortium, verwoordde het recent zo in Science “*Muizen genetica is*

gericht geweest op het eenvoudiger maken van de zaken, wij willen de dingen weer ingewikkelder maken”. Zijn dit nieuwe ideeën? Welnee, de prominente geneticus en statisticus Sir Ronald A. Fisher heeft al begin 1900 de fundamenteën gelegd voor de studie van multifactoriele systemen. Zijn methoden zijn gebaseerd op dieper inzicht, hoe je met een eindige portie geld zo veel mogelijk vragen beantwoord kan krijgen. Waarom wordt er dan toch zo weinig echt multifactorieel geëxperimenteerd? Wellicht vanwege het feit dat de levenswetenschapper de eenvoud van één-factor strategieën prefereert boven de complexiteit van echte multifactoriele methoden. En dit is mogelijk een gevolg van een tekortkoming in de opleiding – hier ligt dus een schone taak voor de bioinformaticus. Opmerkelijk, in de plantenwetenschappen is men minder beducht en viert de kwantitatieve genetika van multifactoriele eigenschappen al jaren hoogtij. In dit opzicht convergeren onderzoek aan plant, dier en mens gelukkig nu weer naar elkaar toe.



Figuur 18. Sir R.A. Fisher (1890-1962).

Om de puzzel op te lossen moeten we dus niet terugschrikken voor complexe zaken, in ieder geval bij studies aan experimentele organismen zoals de muis en *Arabidopsis*. De statistische uitdaging voor vandaag ligt erin dat we de kenmerken van een beperkt aantal individuen willen verklaren en voorspellen aan de hand van vele tienduizenden genen en andere factoren, en een nog groter aantal mogelijke interacties. Dit wordt wel het $n \times P$ probleem genoemd, kleine n voor small number of individuals, en grote P voor large number of predictors. Welke wiskundige, statistische en informatica methoden gebruiken we dan? Kan het überhaupt wel? Het goede nieuws is: het kan wel; het slechte nieuws is: maar lang niet altijd. Om met het ‘niet’ te beginnen: het zal onmogelijk blijven om alle interacties tussen alle genen te schatten, simpelweg omdat we nooit genoeg informatie (muizen, mensen) zullen hebben. Binnen de bioinformatica hoor je vaak klagen over de overdosis aan gegevens (data overload), maar dat is mijns inziens hier dus niet het geval. Het zal voor complexe ziekten



Figuur 19. Opgevoerde (fiets)modellen.

moeilijk blijven om een binaire uitkomst (ziek of gezond) te voorspellen op basis van de RNA-hoeveelheden gemeten aan 30.000 genen bij een beperkt aantal patiënten en gezonde mensen. Wat kan er dan ‘wel’? Zoals je een fiets kunt opvoeren tot grotere prestaties, zo kun je de gangbare één-gen of één-factor modellen opvoeren tot echte multi-gen of multi-factor modellen. De figuur toont een simulatie van een modelsysteem met 5 genen, waarbij het één-gen model op de juiste genoomposities inderdaad kleine piekjes toont, en de multi-gen modellen steeds hogere pieken al naar gelang het model geavanceerder wordt. Waar ligt de grens? Als de meeste genen elkaar gewoon “met rechts de hand schudden” (zoals we eerder hebben gezien), dan kunnen we bij een populatie van 45 muizen de effecten van zo’n 30 genen tegelijkertijd modelleren, bij 100 muizen van zo’n 85 genen, bij 200 muizen van zo’n 180 genen enzovoorts – er blijven dan steeds voldoende vrijheidsgraden over om zinvolle analyses te doen. Is er wel sprake van links- en rechtshandige gen-varianten op twee of meer genen, dan kunnen die interacties gedetecteerd worden op basis van methoden die een beperkt aantal van zulke interacties in de modellen toestaan en op voorwaarde dat de interactie effecten groot genoeg zijn. Deze methoden zijn best ingewikkeld, en het zal een hele uitdaging worden dit in mijn onderwijs en onderzoek zo te vertalen dat de biologen en de medici er de meerwaarde van gaan zien. Voor de

wiskunde en informatica collega's: wij modelleren de effecten van meerdere factoren met behulp van multiële regressie modellen en gegeneraliseerde lineaire modellen, en we gebruiken maximum likelihood, penalized likelihood en Bayesiaanse informatie criteria om de modellen tegen elkaar af te wegen, en bij de berekeningen gebruiken we iteratieve kleinste kwadraten methoden, het EM-algoritme, data imputation en Markov chain Monte Carlo (MCMC) methoden.

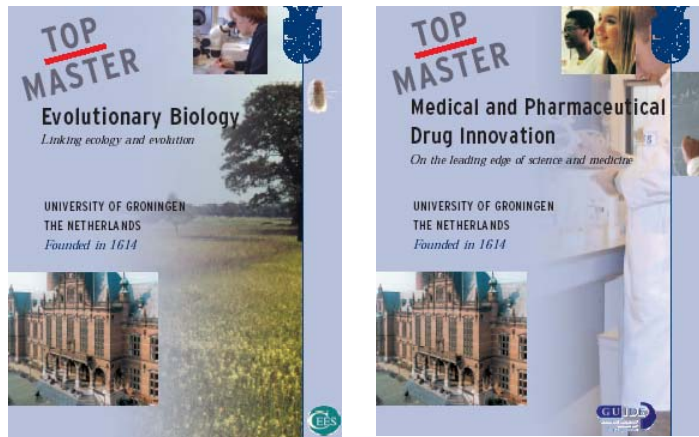
De bioinformatica maakt dus gebruik van een heleboel wiskunde, statistiek en informatica. En dit soort methoden worden niet één keertje gebruikt, maar heel veel keren. Bijvoorbeeld in het 'genetical genomics' project van Gerald de Haan worden op dit moment veel gegevens verzameld: voor elk van 12.000 genen (dus nog niet eens van alle genen) is voor 16 tot 96 deelfragmenten per gen de RNA hoeveelheid gemeten voor 45 muizen. Elk van die vele deelfragmenten dient tevens op kwaliteit gecontroleerd te worden tegen de meest recente genoom informatie van de muis – een flinke rekenklus. Als we dan ook nog permutatie en/of simulatiestudies gaan uitvoeren om significante van toevallige relaties te kunnen onderscheiden, dan loopt het rekenwerk helemaal hard op. Het zal duidelijk zijn, de benodigde rekenintensieve analyses kunnen we nauwelijks meer aan op onze PCs en een samenwerking met het *universitaire rekencentrum* ligt dan ook voor de hand – daar staat ons straks een supercomputer en een cluster van 128 Linux machines ten dienste. Ook de Hanzehogeschool heeft een Linux cluster beschikbaar. Besprekingen met BioASP, de centrale application service provider in de bioinformatica in Nederland, leiden er mogelijk toe dat we in Groningen een mirror-site van BioASP gaan krijgen. Kortom, ook op dit front liggen mooie ontwikkelingen in het verschiet voor Groningen.

Met de juiste wiskunde, statistiek en informatica kunt u beter puzzelen, ik heb u daar net voorbeelden van gegeven. Naar mijn mening mag deze constatering echter niet te vrijblijvend worden opgevat, zeker waar het onderzoek met mensen of proefdieren betreft. Beter puzzelen betekent bijvoorbeeld meer vragen beantwoord krijgen bij gebruik van hetzelfde aantal muizen – door slim en efficiënt multifactorieel te experimenteren. Eerder dit jaar heb ik dit nog eens expliciet onder de aandacht van het lezerspubliek van Nature Reviews Genetics mogen brengen. Beter puzzelen betekent ook: puzzelstukjes van zo veel mogelijk laboratoria combineren – door gegevens nationaal en internationaal meteen en volledig te delen via moderne

informatie-middelen. Zodat we door synergie sneller opschieten met puzzelen. En zodat overbodige herhalingen van experimenten dan achterwege kunnen blijven. Met minder dierenleed als gevolg. Internationale initiatieven, zoals het Complex Trait Consortium, zijn hier expliciet op gericht en verdienen dan ook onze actieve ondersteuning. Ik pleit er voor dat experimenten alleen toegestaan worden als ze aan eerder genoemde voorwaarden voldoen, waarbij aansluiting bij een consortium ook een eis of minimaal een serieus streven zou behoren te zijn. De bioinformatica kan aldus een belangrijke bijdrage leveren aan het terugdringen van het gebruik van proefdieren.

Leren puzzelen

Dames en heren, levensecht puzzelen, dat is waar de bioinformaticus aan bijdraagt. Gezien het belang van de levenswetenschappen, zal het u dan ook niet verwonderen dat diverse universiteiten in de afgelopen jaren een bioinformatica-groep hebben opgestart. Tussen deze groepen zijn er wel duidelijke accentverschillen te herkennen. In Groningen zou je de bioinformatica kunnen typeren als “gericht op de kwantitatieve analyse van complexe eigenschappen”. Dit accent is in mijn ogen niet zo verwonderlijk gezien onze inbedding in, en nauwe band met, de medische en biologische genetica, waar complexe eigenschappen centraal staan. Dit accent zal ook in het onderwijs tot uiting komen, al is het maar omdat ik een groot voorstander ben van onderwijs dat gedreven wordt door de spannende ontwikkelingen bij onze directe samenwerkingspartners – de medische en biologische genetici. Het nieuwe bioinformatica-onderwijs zal zich dus richten op het analyseren van complexe eigenschappen via – en nu volgt een niet uitputtend lijstje – kartering van genen met moleculaire merkers, kwantitatieve analyse op RNA en eiwit niveau, constructie van gen-netwerken, de rol van niet-coderende genoomelementen op expressie van genen, statistische variantieanalysemethoden, informatie-analyse en informatiesystemen, en tenslotte programmeren in diverse programmeertalen en omgevingen. Er zijn hier vast mensen in de zaal die bepaalde onderwerpen missen, zoals bijvoorbeeld de bepaling van de driedimensionale structuur van eiwitten of het dynamisch modelleren van interacties tussen biomoleculaire stoffen met behulp van differentiaal vergelijkingen. Daarop zal het accent in mijn groep inderdaad niet liggen, want in een zo breed veld



Figuur 20. Bioinformatica draagt bij aan de topmasters.

moeten we keuzes maken. Of u dat nu met mij eens bent of niet, u hoeft niet ongerust te worden, want Groningen kent andere groepen die op die terreinen juist beresterk zijn – ik zou zelfs willen zeggen “prof Berendsen-sterk”. Toch zal het onderwijs een sterk interdisciplinair karakter hebben. Ik verwacht in dit kader veel van de nieuwe brede bacheloropleidingen, waarbinnen studenten in de wiskunde en informatica ook een minor in biologie en omics kunnen volgen, en vice versa medici en biologen zich kunnen scholen in wiskunde en informatica. Zodat de studenten daarmee dan voldoende voorgesorteerd zijn voor het volgen van bioinformatica onderwijs. Het nieuwe bachelorcurriculum binnen Life Science and Technology kent dit jaar voor het eerst de hybride keuzevariant genaamd ‘genomics, proteomics and bioinformatics’. Het Groningen Bioinformatica Centrum zal ook graag aan andere opleidingen bijdragen, zo participeren wij bijvoorbeeld al in de topmaster Evolutionary Biology en de topmaster Medical and Pharmaceutical Drug Innovation, en mogelijk later dit jaar in de topmaster Biomolecular Sciences. Ook verzorgen wij onderwijs in de statistiek, van biostatistiek voor biologie en voor Life Science and Technology, tot Mathematical Statistics through Applications voor wiskunde. Er is nog een punt waarmee Groningen zich sterk kan en zal profileren. Wat namelijk niet onvermeld mag blijven, is de nieuwe HBO-studie bioinformatica aan de Hanzehogeschool, waar collega Jan-Peter Nap recent als lector is benoemd. In de praktijk werken WO en HBO bioinformatici vaak nauw in teams samen, en wij – Jan-Peter Nap en ik – beogen dan ook de studenten hier al tijdens de studie ervaring in op te laten doen. De bioinformatica leent zich dus bij uitstek voor zo’n bundeling van universitaire en HBO krachten, met erkenning en benutting van de wezenlijk verschillende rollen van

de HBO-er en de academicus in de beroepspraktijk. Wij bieden daarmee meerwaarde aan de studenten en hun toekomstige werkgevers, en ook aan opdrachtgevers van de studieprojecten. De bioinformatica aan de RuG en aan de Hanzehogeschool zullen zich samen sterk naar buiten profileren onder de naam Groningen Bioinformatica Centrum.

Dames en heren studenten of studenten-in-spé, in het begin van mijn rede heb ik u al verteld dat ons nog heel wat te doen staat. Ik hoop dat ik met deze rede een tipje van de sluier heb kunnen oplichten, zodat u een glimp hebt opgevangen van de vele uitdagingen die nog voor ons liggen. Naar mijn stellige overtuiging bieden de bioinformatica en meer in het algemeen de levenswetenschappen u een reëel perspectief op een mooi beroep, waar u naar harte lust kunt werken aan de grenzen van het weten. U bent van harte welkom om levensecht mee te puzzelen!



Figuur 21. Studenten Life Science and Technology.

Dankwoord

Hiermee ben ik aan het einde van mijn rede gekomen. Mijnheer de Rector, staat u mij toe om hier nog enkele persoonlijke dankwoorden uit te spreken.

Geachte decanen prof Wiersma en prof Poppema, geachte leden van het College van Bestuur en geachte leden van de benoemingsadviescommissie, graag wil ik u bedanken voor mijn gecombineerde benoeming aan de Faculteit der Wiskunde en Natuurwetenschappen en aan de Faculteit der Medische Wetenschappen. Ik beschouw het als een groot voorrecht dat ik Groningen Bioinformatica Centrum vorm mag geven, als een brug tussen deze twee facultaire organisaties, als een brug tussen de facultaire organismen, van mens tot bacterie, van muis tot sluipwesp, en als een brug

tussen diverse disciplines, van wiskunde naar biologie, van informatica tot geneeskunde.

In het afgelopen jaar is Groningen Bioinformatica Centrum snel gegroeid, onder andere ten gevolge van het binnenhalen van een aantal subsidies. Gerard te Meerman, Peter Terpstra, Ilja Nolte, Menno Hardonk, Edo Plantenga, Casper Albers, Stephen Nyangoma, Morris Swertz, Rudi Alberts, Jingyuan Fu, Martijn Dijkstra, Bert de Brock, James Pass, Jan-Peter Nap, Desiree Hansen, en binnenkort Theo Reijmers. Beste collega's, ik wil jullie namen graag in het openbaar genoemd hebben, en ik wil laten weten hoe ik onder de indruk ben van de wijze waarop we nu al met de nodige teamgeest met elkaar werken. Laten we zo door gaan!

Groningen Bioinformatica Centrum heeft met vele groepen binnen de RuG een samenwerkingsrelatie. De biologische en biomedische vraagstellingen van deze partners leveren ons de vruchtbare voedingsbodem waarop wij als Groningen Bioinformatica Centrum kunnen groeien. Hoewel ik besef dat ik daarmee anderen te kort doe, zal ik enkele groepen met name noemen: medische genetica van prof Charles Buys, genetica van prof Oscar Kuipers en prof Leo Beukeboom bij biologie, stamcelbiologie van dr Gerald de Haan, het genotyperingslab bij pathologie van dr Gerrit van der Steege, de proteomics laboratoria van prof Rainer Bischoff bij farmacie en prof Roel Vonk bij kindergeneeskunde, en het breedtestrategieproject onderleiding van prof Gerjan Navis van interne geneeskunde.

Mijn wetenschappelijke loopbaan is gevormd door interactie met vele collega's. Om te beginnen wil ik Willem Schaafsma bedanken voor de originele en inspirerende wijze waarop hij mij de mathematische statistiek heeft bijgebracht. Na mijn studie in Groningen ben ik in 1988 naar Wageningen vertrokken. Daar is Hans Jansen jarenlang mijn directe collega en chef geweest en hij heeft mij perfect wegwijs gemaakt in de wereld van de statistiek en biometrie. In 1991 kwam ook prof Piet Stam ten tonele en hij bracht samen met prof Maarten Koornneef de kwantitatief genetische vraagstellingen binnen mijn vizier. Ik profiteer hiervan nog tot aan de dag van vandaag. Piet werd hoogleraar en ik kon mooi meteen als eerste bij hem promoveren in 1995.

Vanaf 1995 heb ik met twee moleculaire biologen veel mogen samenwerken. Met Hans Sandbrink heb ik een jaar lang de werkkamer gedeeld, om aldus doordrenkt te raken in de biologische vraagstellingen en ontwikkelingen. Helaas, Hans kreeg hartproblemen en is hij iets minder dan twee jaar geleden veel te jong overleden – dat doet nog steeds pijn. Ik heb mijn Nature Reviews Genetics paper van februari dit jaar dan ook aan hem opgedragen. De andere moleculaire bioloog, aan wie ik heel veel te danken heb, is Jan-Peter Nap. De samenwerking met hem is een toonbeeld van hoe leuk, inspirerend en vruchtbaar multidisciplinair samenwerken kan zijn. Ik ben dan ook heel blij dat ook hij de weg naar het noorden heeft gevonden, als lector aan de Hanzehogeschool en heel binnenkort als associate professor binnen Groningen Bioinformatica Centrum met de leeropdracht Toegepaste Bioinformatica.

Mijn bijzondere dank gaat ook uit naar mijn vader Gerard en moeder Netty. Ik kom uit een biologisch en academisch georiënteerd nest. Mijn vader heeft mij in 1981 gestimuleerd om wiskunde te gaan studeren en vandaar uit de biologie als toepassingsgebied te veroveren, en aldus geschiedde. Door de jaren heen heeft mijn vader altijd mijn artikelen-in-wording willen lezen en bekritisieren, ik heb heel veel van hem geleerd. Pa, zonder twijfel sta ik hier omdat jij je ermee hebt bemoeid.

Een speciaal dankwoord is er voor mijn gezin. Lieve Henny, wij hebben samen de afgelopen 20 jaar veel beleefd en zijn productief geweest op het werk en thuis. Ik ben je heel dankbaar voor de vrijheid en enorme ondersteuning die je mij hebt gegeven in alle fases van student tot waar ik nu sta. Mensen helpen zich te ontwikkelen, daar ben je nu ook beroepshalve als coach actief mee. We hebben ook ons eigen genetische experiment opgezet, want we kregen drie kinderen. Lieve Rianne, Ymke en Yde, ik vind het heel bijzonder dat ik vandaag deze spreekbeurt voor jullie heb mogen houden!

Levensecht puzzelen, wel, daar gaan we morgen maar mee verder, want voor vandaag zijn ben ik in ieder geval uitgepuzzeld. Ik dank u allen hartelijk voor uw komst en uw aandacht. Het is tijd voor de receptie!

Ik heb gezegd

Geselecteerde publicaties 2001-2003 (auteurs Groningen Bioinformatica Centrum in cursief)

Complexe eigenschappen bij de mens:

1. Fischer, C., Beckmann, L., Majoram, P., *Te Meerman, G.J.*, Chang-Claude, J. (2003) Haplotype sharing analysis with SNPs in candidate genes: the Genetic Analysis Workshop 12 example. **Genetic Epidemiology** 24: 68-73
2. Jong, M.M. de, *Nolte, I.M.*, *Te Meerman, G.J.*, Graaf, W.T. van der, Vries, E.G. de, *Sijmons R.H.*, Hofstra, R.M., Kleibeuker J.H. (2002) Low-penetrance genes and their involvement in colorectal cancer susceptibility. **Cancer Epidemiology Biomarker** 11: 1332-52
3. Jong, M.M. de, *Nolte I.M.*, *Te Meerman, G.J.*, Graaf, W.T. de, Oosterwijk, J.C., Kleibeuker, J.H., Schaapveld, M., Vries, E.G. de. (2002) Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility. **Journal of Medical Genetics** 39: 225-42
4. *Nolte, I.M.*, *Te Meerman G.J.* (2002) The probability that similar haplotypes are identical by descent. **Annals of Human Genetics** 66: 195-209
5. *Te Meerman, G.J.*, Vries, E.G. de. Relevance of high and low penetrance.(2001) **Lancet** 28;358 (9278): 331-2
6. Boon, M., *Nolte, I.M.*, Bruinenberg, M., Spijker, G.T., Terpstra, P., Raelson, J., Keyser, J., Zwanikken, C.P., Hulsbeek, M., Hofstra, R.M., Buys, C.H., *Te Meerman, G.J.* (2001) Mapping of a susceptibility gene for multiple sclerosis to the 51 kb interval between G511525 and D6S1666 using a new method of haplotype sharing analysis. **Neurogenetics** 3: 221-30

Complexe eigenschappen bij modelsystemen:

7. Complex Trait Consortium (incl *Jansen, R.C.*) (2003) The nature and identification of quantitative trait loci: a community's view. **Nature Reviews Genetics** 4: 911–916
8. *Jansen, R.C.*, J.L. Jannink, W.D. Beavis (2003) Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. **Crop Science** 43: 829–834.

9. Boer, M.P., C.J.F. ter Braak, *R.C. Jansen* (2002). A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome scans. **Genetics** 162: 951-960
10. Jannink, J.-L., M.C.A.M. Bink and *R.C. Jansen* (2001) Using complex plant pedigrees to map valuable genes. **Trends in Plant Science** 6(8): 337-342
11. Jannink, J.-L. and *R.C. Jansen* (2001) Mapping epistatic QTL with one-dimensional genome searches. **Genetics** 157: 445-454

Genetical genomics:

12. *Jansen, R.C.* (2003) Studying complex biological systems using multifactorial perturbation. **Nature Reviews Genetics** 4: 145-151
13. Yvert, G. *et al.* (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. **Nature Genetics** 35: 57-64
14. Darvasi, A. (2003) Genomics: gene expression meets genetics. **Nature** 422: 269-270
15. Vogel, G. (2003) Scientists dream of 1001 complex mice. **Science** 301: 456-457
16. Schadt, E.E. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. **Nature**, 422: 297-302
17. Brem, R.B. *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. **Science** 296: 752-755
18. Klose, J. *et al.* (2002) Genetic analysis of the mouse brain proteome. **Nature Genetics** 30: 385-393
19. *Jansen, R.C.* and *J.P. Nap* (2001) Genetical genomics: the added value from segregation. **Trends in Genetics** 17: 388-391

Divers:

20. Mlynarova, L., A. Loonen, E. Mietkiewska, *R.C. Jansen* and *J.P. Nap* (2002) Assembly of two transgenes in an artificial chromatin domain gives highly co-ordinated expression in tobacco. **Genetics** 160:727-740
21. *Jansen, R.C.*, L. Mlynarova and *J.P. Nap*. (2002) Errors in genomics and proteomics. **Nature Biotechnology** 20(10): 19
22. *Jansen, R.C.*, H. Geerlings, A.J. van Oeveren and R.C. van Schaik (2001) A comment on co-dominant scoring of AFLP markers. **Genetics** 158: 925-926